

DOCKET NO.: 261119US90PCT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Kazunori KOBAYASHI, et al.

SERIAL NO.: NEW U.S. PCT APPLICATION

FILED: HEREWITH

INTERNATIONAL APPLICATION NO.: PCT/JP04/01261

INTERNATIONAL FILING DATE: February 6, 2004

FOR: SOUND ACQUISITION METHOD AND SOUND ACQUISITION APPARATUS

REQUEST FOR PRIORITY UNDER 35 U.S.C. 119
AND THE INTERNATIONAL CONVENTIONCommissioner for Patents
Alexandria, Virginia 22313

Sir:

In the matter of the above-identified application for patent, notice is hereby given that the applicant claims as priority:

<u>COUNTRY</u>	<u>APPLICATION NO</u>	<u>DAY/MONTH/YEAR</u>
Japan	2003-030676	07 February 2003
Japan	2003-058626	05 March 2003

Certified copies of the corresponding Convention application(s) were submitted to the International Bureau in PCT Application No. PCT/JP04/01261. Receipt of the certified copy(s) by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

Respectfully submitted,
OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



Masayasu Mori
Attorney of Record
Registration No. 47,301
Surinder Sachar
Registration No. 34,423

Customer Number

22850

(703) 413-3000
Fax No. (703) 413-2220
(OSMMN 08/03)

Corwin P. Umbach, Ph.D.
Registration No. 40,211

Rec'd PCT/PTO 29 OCT 2004

10/510955

T/JP 2004/001261

日本国特許庁
JAPAN PATENT OFFICE

06.2.2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年 3月 5日

出願番号
Application Number: 特願2003-058626

[ST. 10/C]: [JP2003-058626]

出願人
Applicant(s): 日本電信電話株式会社

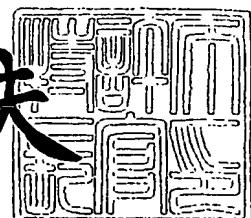
RECEIVED
25 MAR 2004
WIPO PCT

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

2004年 3月 12日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



BEST AVAILABLE COPY

出証番号 出証特2004-3019562

【書類名】 特許願

【整理番号】 NTTI147129

【提出日】 平成15年 3月 5日

【あて先】 特許庁長官 殿

【国際特許分類】 G01L 15/28

【発明者】

【住所又は居所】 東京都千代田区大手町二丁目 3番 1号 日本電信電話株式会社内

【氏名】 小林 和則

【発明者】

【住所又は居所】 東京都千代田区大手町二丁目 3番 1号 日本電信電話株式会社内

【氏名】 古家 賢一

【特許出願人】

【識別番号】 000004226

【氏名又は名称】 日本電信電話株式会社

【代理人】

【識別番号】 100123788

【弁理士】

【氏名又は名称】 宮崎 昭夫

【電話番号】 03-3585-1882

【選任した代理人】

【識別番号】 100106297

【弁理士】

【氏名又は名称】 伊藤 克博

【手数料の表示】

【予納台帳番号】 201087

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 収音方法、装置、およびプログラム

【特許請求の範囲】

【請求項1】 収音方法であつて、

複数のマイクロホンの各々で受音されたマイクロホン受音信号から発話区間を判定する発話判定段階と、

前記発話判定段階において発話区間と判定された場合に、前記マイクロホン受音信号から音源位置を検出する音源位置検出段階と、

前記マイクロホン受音信号から共分散行列を算出する共分散行列算出段階と、

前記音源位置検出段階の検出結果に基づき、前記共分散行列を音源位置ごとに記憶する共分散行列記憶段階と、

前記各音源位置に対する共分散行列に基づき、音源位置ごとの収音レベルを求める収音レベル推定段階と、

前記共分散行列および前記音源位置ごとの収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるような複数チャネルのフィルタ係数および増幅率を算出するフィルタ係数算出段階と、

前記マイクロホン受音信号を、前記複数チャネルのフィルタ係数で各々フィルタリングする複数のフィルタ段階と、

前記複数のフィルタ段階の出力信号を加算する加算段階と、

前記加算段階の出力信号を前記増幅率で増幅する可変増幅段階とを有する収音方法。

【請求項2】 前記発話判定段階の代りに、複数のマイクロホンの各々で受音されたマイクロホン受音信号から発話区間および雑音区間を判定する発話・雑音判定段階を有し、前記音源位置検出段階の代りに、前記発話・雑音判定段階において発話区間と判定された場合に、前記マイクロホン受音信号から音源位置を検出する音源位置検出段階を有し、前記共分散行列記憶段階の代りに、前記送受話判定段階の判定結果および前記音源位置検出段階の検出結果に基づき、前記共分散行列を雑音区間と音源位置ごとに記憶する共分散行列記憶段階を有する、請求項1に記載の収音方法。

【請求項3】 前記発話判定段階の代りに、複数のマイクロホンの各々で受音されたマイクロホン受音信号と受話信号から送話区間、受話区間、および雑音区間を判定する送受話判定段階を有し、前記音源位置検出段階の代りに、前記送受話判定段階において発話区間と判定された場合に、前記マイクロホン受音信号から音源位置を検出する音源位置検出段階を有し、前記共分散行列算出段階の代りに、前記マイクロホン受音信号および前記受話信号から共分散行列を算出する共分散行列算出段階を有し、前記共分散行列記憶段階の代りに、前記送受話判定段階の判定結果および前記音源位置検出段階の検出結果に基づき、前記共分散行列を雑音区間と、受話区間と、音源位置ごとに記憶する共分散行列記憶段階を有する、請求項1に記載の収音方法。

【請求項4】 前記音源位置に対する感度拘束の重み $C_{S1} \sim C_{Sn}$ を、過去に発音した音源に対する重みほど減少させる、請求項1から請求項3のいずれかに記載の収音方法。

【請求項5】 前記フィルタ係数算出段階が、各々の共分散行列

【外1】

$$R_{xx}(\omega)$$

の対角成分

【外2】

$$\text{diag}(R_{xx}(\omega))$$

と任意のM行またはM+1行の行列

【外3】

$$D$$

に基づく重み

【外4】

$$1/\{D^H \text{diag}(R_{xx}(\omega)) D\}$$

を各々の共分散行列

【外5】

$$R_{xx}(\omega)$$

に乘じ白色化してから、フィルタ係数の算出をする、請求項1から請求項4のいずれかに記載の収音方法。

【請求項6】 前記共分散行列記憶段階が、過去に記憶されている共分散行列と、前記共分散行列算出部により新たに算出された共分散行列を平均して、現在の共分散行列として記憶する、請求項1から請求項5のいずれかに記載の収音方法。

【請求項7】 収音装置であって、

複数のマイクロホンの各々で受音されたマイクロホン受音信号から発話区間を判定する発話判定手段と、

前記発話判定手段において発話区間と判定された場合に、前記マイクロホン受音信号から音源位置を検出する音源位置検出手段と、

前記マイクロホン受音信号から共分散行列を算出する共分散行列算出手段と、

前記音源位置検出手段の検出結果に基づき、前記共分散行列を音源位置ごとに記憶する共分散行列記憶手段と、

前記各音源位置に対する共分散行列に基づき、音源位置ごとの収音レベルを求める収音レベル推定手段と、

前記共分散行列および前記音源位置ごとの収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるような複数チャネルのフィルタ係数および増幅率を算出するフィルタ係数算出手段と、

前記マイクロホン受音信号を、前記複数チャネルのフィルタ係数で、各々フィルタリングする複数のフィルタ手段と、

前記複数のフィルタ手段の出力信号を加算する加算手段と、

前記加算手段の出力信号を前記増幅率で増幅する可変増幅手段とを有する収音装置。

【請求項8】 請求項1から6のいずれかに記載の収音方法をコンピュータ

に実行させるための収音プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は収音方法および装置に関する。

【0002】

【従来の技術】

図5に従来の収音装置のブロック図を示す。

【0003】

この従来の収音装置はマイクロホン301とパワー算出部302と増幅率設定部303と増幅器304により構成される（例えば、特許文献1の図4、p2参照）。

【0004】

パワー算出部302は、マイクロホン301で受音された信号の長時間平均パワー P_{ave} を求める。長時間平均パワーは、受音信号を二乗し、時間積分することで求めることができる。次に、増幅率設定部303は、パワー算出部302で算出された受音信号の長時間平均パワー P_{ave} と、あらかじめ設定した所望の送出レベル P_{opt} に基づき増幅率Aを設定する。増幅率Aは、例えば、式（1）により求められる。

【0005】

【数1】

$$A = \sqrt{\frac{P_{opt}}{P_{ave}}} \quad (1)$$

【0006】

増幅器304は、設定された増幅率Aで、マイクロホン受音信号を増幅し、出力する。

【0007】

以上の処理により、出力信号パワーは、所望の送出レベル P_{opt} となり、音量

が自動的に調整される。

【0008】

【特許文献1】

特開平8-250944号公報

【0009】

【発明が解決しようとする課題】

しかし、従来の収音装置では、長時間平均パワーを基に増幅率を決定しているので、適切な増幅率が設定されるまでに数秒から数十秒の遅延が生じる。これにより、話者が複数人存在し、それぞれの話者音声が異なるレベルでマイクロホンに収音される場合、発話者が切り替るたびに、増幅率設定の適応が遅れ、不適切な音量となるという問題が生じる。

【0010】

本発明の目的は、話者が複数人存在し、それぞれの話者音声が異なるレベルで収音手段に収音される場合においても、各話者音声が適切な音量となるように音量を自動調整する収音方法、装置、およびプログラムを提供することである。

【0011】

【課題を解決するための手段】

上記目的を達成するために、本発明の収音装置は、

複数のマイクロホンの各々で受音されたマイクロホン受音信号から発話区間を判定する発話判定手段と、

発話判定手段において発話区間と判定された場合に、マイクロホン受音信号から音源位置を検出する音源位置検出手段と、

マイクロホン受音信号から共分散行列を算出する共分散行列算出手段と、

音源位置検出手段の検出結果に基づき、共分散行列を音源位置ごとに記憶する共分散行列記憶手段と、

各音源位置に対する共分散行列に基づき、音源位置ごとの収音レベルを求める収音レベル推定手段と、

共分散行列および前記音源位置ごとの収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるような複数チ

ヤネルのフィルタ係数および増幅率を算出するフィルタ係数算出手段と、
マイクロホン受音信号を、複数チャネルのフィルタ係数で、各々フィルタリングする複数のフィルタ手段と、
複数のフィルタ手段の出力信号を加算する加算手段と、
加算手段の出力信号を前記増幅率で増幅する可変増幅手段とを有する。

【0012】

本発明は、共分散行列および音源位置ごとの収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるような複数チャネルのフィルタ係数と増幅率を算出することにより、話者が複数人存在し、それぞれの話者音声が異なるレベルでマイクロホンに収音される場合においても、複数のマイクロホンを用いて指向特性を適切に制御することにより、話者ごとに適切な音量となるように音量を自動調整する。

【0013】

【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して説明する。

【0014】

(第1の実施形態)

図1は本発明の第1の実施形態の収音装置のブロック図である。

【0015】

本実施形態の収音装置は、マイクロホン1021～102Mとフィルタ1031～103Mと加算器104と発話判定部105と音源位置検出部106と共に分散行列算出部107と共に分散行列記憶部108と収音レベル推定部109とフィルタ係数算出部110と可変増幅器111により構成される。

【0016】

本実施形態は、音源位置を推定し、各音源位置に対する共分散行列を保存しておき、これらを用いてフィルタ係数を求める。これらのフィルタ係数で、マイクロホン受音信号をフィルタリングすることで、各音源からの信号が一定の音量となるように制御する。

【0017】

まず、発話判定部105では、マイクロホン102₁～102_Mで受音したマイクロホン受音信号から、発話区間を検出する。例えば、マイクロホン受音信号について、短時間平均パワー（0.1～1s程度）と長時間平均パワー（1s～100s程度）を求め、短時間平均パワーと長時間平均パワーの比が発話の閾値を超える場合に発話区間と判定する。

【0018】

発話判定部105の判定結果が発話区間であった場合に、音源位置検出部106は、音源の位置を推定する。音源位置の推定方法は、例えば相互相関法による方法がある。

【0019】

M個のマイクロホンがあると想定し、i番目のマイクロホン102_iとj番目のマイクロホン102_jで受音された信号より求められる受音信号間遅延時間差を τ_{ij} とする。受音信号間遅延時間差は、信号間の相互相関を求め、その最大ピーク位置から求めることができる。次に、m番目の受音位置を（ x_m, y_m, z_m ）、推定音源位置を

【0020】

【外6】

$$(\hat{X}, \hat{Y}, \hat{Z})$$

【0021】

と表す。これらの位置から求められる推定受音信号間遅延時間差

【0022】

【外7】

$$\hat{\tau}_{ij}$$

【0023】

は式（2）で表される。

【0024】

【数2】

$$\hat{\tau}_y = \frac{1}{c} \cdot \sqrt{(x_i - \hat{X})^2 + (y_i - \hat{Y})^2 + (z_i - \hat{Z})^2} - \frac{1}{c} \cdot \sqrt{(x_j - \hat{X})^2 + (y_j - \hat{Y})^2 + (z_j - \hat{Z})^2} \quad (2)$$

ただし、cは音速

【0025】

次に、受音信号間遅延時間差

【0026】

【外8】

 τ_{ij} 、 $\hat{\tau}_{ij}$

【0027】

に音速cを乗じ距離に換算したものを、それぞれ受音位置間距離差

【0028】

【外9】

 d_{ijk} 、 d_{ij}

【0029】

とし、測定値 d_{ij} と推定値

【0030】

【外10】

 \hat{d}_{ij}

【0031】

の二乗平均誤差

【0032】

【外11】

 $e(\mathbf{q})$

【0033】

を求めれば、式（3）となる。

【0034】

【数3】

$$e(\mathbf{q}) = \sum_{i=1}^{M-1} \sum_{\substack{j=1 \\ j \neq i}}^M \left| d_{ij} - \hat{d}_{ij} \right|^2$$

$$= \sum_{i=1}^{M-1} \sum_{\substack{j=1 \\ j \neq i}}^M \left| d_{ij} - \sqrt{(x_i - \hat{x})^2 + (y_i - \hat{y})^2 + (z_i - \hat{z})^2} + \sqrt{(x_i - \hat{x})^2 + (y_i - \hat{y})^2 + (z_i - \hat{z})^2} \right|^2 \quad (3)$$

ただし、 $\mathbf{q} = (\hat{X}, \hat{Y}, \hat{Z})$ である。

【0035】

式（3）の二乗平均誤差

【0036】

【外12】

$e(\mathbf{q})$

【0037】

を最小化する解を求めるには、受音信号間遅延時間差の測定値と推定値の誤差が最小となる推定音源位置を求めることができる。ただし、式（3）は非線形連立方程式となっており、解析的に解くことは困難であるので、逐次修正を用いた数値解析により求める。

【0038】

式（3）を最小化する推定音源位置

【0039】

【外13】

$(\hat{X}, \hat{Y}, \hat{Z})$

【0040】

を求めるには、ある点における勾配を求め、誤差が小さくなる方向に推定音源位置を修正していく、勾配が0となる点を求めるので、修正式は式（4）の

ようになる。

【0041】

【数4】

$$\mathbf{q}_{(u+1)} = \mathbf{q}_{(u)} - \alpha \cdot \text{grad } e(\mathbf{q}) \Big|_{\mathbf{q}=\mathbf{q}_{(u)}} \quad (4)$$

ただし、 α は修正のステップサイズ、 $\mathbf{q}_{(u)}$ は、 u 回修正後の \mathbf{q} を表し、 grad は勾配を表し、式(5)～式(9)で表される。

$$\text{grad } e(\mathbf{q}) = \left(\frac{\partial e(\mathbf{q})}{\partial \hat{X}}, \frac{\partial e(\mathbf{q})}{\partial \hat{Y}}, \frac{\partial e(\mathbf{q})}{\partial \hat{Z}} \right) \quad (5)$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{X}} = 2 \sum_{i=1}^{M-1} \sum_{\substack{j=1 \\ j>i}}^M \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{x_i - \hat{X}}{r_i} - \frac{x_j - \hat{X}}{r_j} \right\} \quad (6)$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{Y}} = 2 \sum_{i=1}^{M-1} \sum_{\substack{j=1 \\ j>i}}^M \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{y_i - \hat{Y}}{r_i} - \frac{y_j - \hat{Y}}{r_j} \right\} \quad (7)$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{Z}} = 2 \sum_{i=1}^{M-1} \sum_{\substack{j=1 \\ j>i}}^M \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{z_i - \hat{Z}}{r_i} - \frac{z_j - \hat{Z}}{r_j} \right\} \quad (8)$$

$$r_m = \sqrt{(x_m - \hat{X})^2 + (y_m - \hat{Y})^2 + (z_m - \hat{Z})^2} \quad (9)$$

【0042】

以上、式(4)を繰返し計算することで、誤差が最小となる推定音源位置を求めることができる。

【0043】

次に、共分散行列算出部107では、マイクロホン受音信号の共分散を求め、それを行列にする。まず、マイクロホン受音信号の周波数領域変換信号を $X_1(\omega) \sim X_M(\omega)$ とする。これらの信号の共分散行列

【0044】

【外14】

$$R_{xx}(\omega)$$

【0045】

は、式(10)により算出される。

【0046】

【数5】

$$R_{xx}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_M(\omega) \end{pmatrix} \begin{pmatrix} X_1(\omega)^* & \cdots & X_M(\omega)^* \end{pmatrix} \quad (10)$$

ただし、*は共役を表す。

【0047】

次に、共分散行列記憶部108では、音源位置検出部106の検出結果に基づき、共分散行列

【0048】

【外15】

$$R_{xx}(\omega)$$

【0049】

を、各音源位置に対する共分散行列

【0050】

【外16】

$$R_{s_i s_i}(\omega)$$

【0051】

として保存する。

【0052】

収音レベル推定部109は、各音源の収音レベル P_{Si} を、各音源位置に対する共分散行列

【0053】

【外17】

$$\mathbf{R}_{S1S1}(\omega) \sim \mathbf{R}_{SNSN}(\omega)$$

【0054】

と、各音源位置ごとの任意のM行の行列

【0055】

【外18】

$$\mathbf{G}_1(\omega) \sim \mathbf{G}_n(\omega)$$

【0056】

に基づく式(11)により求める。

【0057】

【外19】

$$\mathbf{G}_i(\omega)$$

【0058】

は、例えば、1番目のマイクロホン1021の信号のみを取り出す行列

【0059】

【外20】

$$\mathbf{G}_i(\omega) = (1 \ 0 \ \cdots \ 0)$$

【0060】

などである。

【0061】

【数6】

$$P_{Si} = \frac{1}{W} \sum_{\omega=0}^W \mathbf{G}_i(\omega)^H \mathbf{R}_{SiS_i}(\omega) \mathbf{G}_i(\omega) \quad (11)$$

【0062】

次に、フィルタ係数算出部110では、各音源から発せられた音を所望の音量で収音するためのフィルタ係数を計算する。まず、各マイクロホンに接続されたフィルタ103₁～103_Mのフィルタ係数を周波数領域に変換したものをH₁(ω)～H_M(ω)とする。次に、これらのフィルタ係数を式(12)により行列としたものを

【0063】

【外21】

 $\mathbf{H}(\omega)$

【0064】

とする。

【0065】

【数7】

$$\mathbf{H}(\omega) = \begin{pmatrix} H_1(\omega) \\ \vdots \\ H_M(\omega) \end{pmatrix} \quad (12)$$

【0066】

また、i番目の音源が発音している期間のマイクロホン受音信号の周波数領域変換信号をX_{Si,1}(ω)～X_{Si,M}(ω)とする。

【0067】

ここで、フィルタ係数行列

【0068】

【外22】

 $H(\omega)$

【0069】

に要求される条件は、マイクロホン受音信号をフィルタ係数行列

【0070】

【外23】

 $H(\omega)$

【0071】

でそれぞれフィルタリングし、フィルタリング後の信号を加算したときに、各音源の信号成分のレベルが所望のレベル P_{opt} となっていることである。したがって、各音源の信号をフィルタリングおよび加算した信号が、マイクロホン受音信号に任意のM行の行列

【0072】

【外24】

 $G_s(\omega)$

【0073】

と所望の利得を乗じた信号となる式(13)が理想条件となる。

【0074】

【数8】

$$(X_{s,1}(\omega) \ \cdots \ X_{s,M}(\omega))H(\omega) = \sqrt{\frac{P_{opt}}{P_s}} (X_{s,1}(\omega) \ \cdots \ X_{s,M}(\omega))G_s(\omega) \quad (13)$$

【0075】

次に、式(13)の条件をフィルタ係数行列

【0076】

【外25】

 $H(\omega)$

【0077】

について最小二乗解で解けば、式（14）となる。ただし、 C_{Si} は音源位置の感度拘束に対する重みの定数であり、値が大きくなるほど感度拘束が強くなる。

【0078】

【数9】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) \right\}^{-1} \sum_{i=1}^n C_{Si} \sqrt{\frac{P_{opt}}{P_{Si}}} R_{SiSi}(\omega) G_i(\omega) \quad (14)$$

【0079】

以上で、音量を調整するためのフィルタ係数を求める式（14）を導出した。

【0080】

次に、式（14）により求められたフィルタ係数は、フィルタ部1031～103Mにコピーされ、マイクロホン受音信号をそれぞれフィルタリングする。フィルタリング後の信号は加算器104で加算され、可変増幅器111においてあらかじめ設定された固定増幅率で増幅された後、出力信号として出力される。

【0081】

また、式（14）を式（15）に置き換えてフィルタ係数を算出し、可変増幅器111の増幅率Bを最も新しく発音のあった音源の収音レベル P_{snew} から式（16）により設定するという方法もある。

【0082】

【数10】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) \right\}^{-1} \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) G_i(\omega) \quad (15)$$

$$B = \sqrt{\frac{P_{opt}}{P_{snew}}} \quad (16)$$

【0083】

次に、本発明の有効性を示すために、図4にシミュレーション結果を示す。シミュレーション条件は、マイクロホン数M：4、信号対雑音比：約20dB、部屋の残響時間：300ms、話者数：2（話者Aはマイクロホンから50cmの位置、話者Bはマイクロホンから200cmの位置）である。この条件で、話者Aと話者Bが交互に発音した場合のマイクロホン受音信号波形を図5（a）に示す。話者Aと話者Bの音声波形を比較すると、話者Bの音声の振幅が低くなっていることが分かる。これに対し、本発明により処理を行った後の波形を図5（b）に示す。話者Aと話者Bの音声波形の振幅の大きさがほぼ同じとなっており、音量調整の効果が確認できる。

【0084】

以上示したように、本実施形態では、複数マイクロホンの受音信号から、音源位置を推定し、共分散行列を求め、音源位置ごとに音量を調整するフィルタ係数を求め、それらのフィルタ係数でマイクロホン受音信号をフィルタリングすることで、話者位置ごとに音量を調整した出力信号を得ることができる。

【0085】

（第2の実施形態）

図2は本発明の第2の実施形態である収音装置のブロック図である。

【0086】

本実施形態の収音装置はマイクロホン1021～102Mとフィルタ1031～103Mと加算器104と発話・雑音判定部201と音源位置検出部106と共に分散行列算出部107と共に分散行列記憶部108と収音レベル推定部109とフィルタ係数算出部110と可変増幅器111により構成される。

【0087】

本実施形態は、第1の実施形態の収音装置の音量調整に雑音抑圧を加えたものである。

【0088】

まず、発話・雑音判定部201では、マイクロホン1021～102Mで受音したマイクロホン受音信号のパワーから発話区間および雑音区間を検出する。例え

ば、それぞれのマイクロホン受音信号について、短時間平均パワー（0.1～1s程度）と長時間平均パワー（1s～100s程度）を求め、短時間平均パワーと長時間平均パワーの比が発話の閾値を超えた場合に発話区間と判定し、雑音の閾値未満の場合に雑音区間と判定する。

【0089】

発話・雑音判定部201の判定結果が発話区間であった場合に、音源位置検出部106は、第1の実施形態と同様にして音源位置を検出する。

【0090】

次に、共分散行列算出部107は、第1の実施形態と同様にして、共分散行列

【0091】

【外26】

$$R_{xx}(\omega)$$

【0092】

を算出する。

【0093】

共分散行列記憶部108では、音源位置検出部106の検出結果および発話・雑音判定部201の判定結果に基づき、共分散行列

【0094】

【外27】

$$R_{xx}(\omega)$$

【0095】

を、各音源位置に対する共分散行列

【0096】

【外28】

$$R_{s_i s_i}(\omega)$$

【0097】

と雑音区間の共分散行列

【0098】

【外29】

$\mathbf{R}_{NN}(\omega)$

【0099】

として保存する。

【0100】

収音レベル推定部109は、各音源の収音レベル P_{Si} を、第1の実施形態と同様にして推定する。

【0101】

次に、フィルタ係数算出部110では、各音源から発せられた音を所望の音量で収音し、雑音を抑圧するためのフィルタ係数を算出する。まず、雑音抑圧の条件を求める。雑音区間のマイクロホン受音信号の周波数領域変換信号を $X_{N,1}(\omega) \sim X_{N,M}(\omega)$ とする。雑音区間のマイクロホン受音信号 $X_{N,1}(\omega) \sim X_{N,M}(\omega)$ がフィルタ103₁～103_Mおよび加算器104通過後に0となれば、雑音を抑圧できているので、雑音抑圧のための条件式は、式(17)となる。

【0102】

【数11】

$$(X_{N,1}(\omega) \ \cdots \ X_{N,M}(\omega)) \mathbf{H}(\omega) = 0 \quad (17)$$

【0103】

この式(17)と、第1の実施形態で示した音量調整のための条件式(13)の両方を満たせば、音量調整と雑音抑圧の両方を実現できる。

【0104】

次に、式(13) (14)、(17)の条件をフィルタ係数行列

【0105】

【外30】

$$H(\omega)$$

【0106】

について最小二乗解で解けば、式（18）となる。ただし、 C_N は雑音消去量に対する重みの定数であり、値が大きくなるほど雑音の消去量が増加する。その他の記号の意味については、第1の実施形態と同様である。

【0107】

【数12】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) + C_N R_{NN}(\omega) \right\}^{-1} \sum_{i=1}^n C_{Si} \sqrt{\frac{P_{Si}}{P_{Si}}} R_{SiSi}(\omega) G_i(\omega) \quad (18)$$

【0108】

以上で、音量を調整し、雑音を抑圧するためのフィルタ係数を求める式（18）を導出した。

【0109】

次に、式（18）により求められたフィルタ係数はフィルタ部1031～1033にコピーされ、マイクロホン受音信号をそれぞれフィルタリングする。フィルタリング後の信号は、加算器104で加算され、可変増幅器111においてあらかじめ設定した固定増幅率で増幅された後、出力信号として出力される。

【0110】

また、式（18）を式（19）に置き換えてフィルタ係数を算出し、可変増幅器111の増幅率Bを、第1の実施形態と同様に最も新しく発音のあった音源の収音レベル P_{Snew} から式（16）により設定するという方法もある。

【0111】

【数13】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) + C_N R_{NN}(\omega) \right\}^{-1} \sum_{i=1}^n C_{Si} R_{SiSi}(\omega) G_i(\omega) \quad (19)$$

【0112】

以上、第2の実施形態は、第1の実施形態における音量調整の効果に加え、雑音抑圧を行うことが可能である。

【0113】

これら以外の部分に関しては、第1の実施形態と同じであるので、説明を省略する。

【0114】

(第3の実施形態)

図3は、本発明の第3の実施形態の収音装置のブロック図である。

【0115】

本実施形態の収音装置はスピーカ301とマイクロホン102₁～102_Mとフィルタ103₁～103_M、302と加算器104と送受話判定部303と音源位置検出部106と共に分散行列算出部107と共に分散行列記憶部108と収音レベル推定部109とフィルタ係数算出部110と可変増幅器111により構成される。

【0116】

本実施形態は、第2の実施形態の収音装置の音量調整および雑音抑圧にエコー抑圧を加えたものである。

【0117】

まず、送受話判定部303では、マイクロホン102₁～102_Mで受音したマイクロホン受音信号および受話信号のパワーから送話区間、受話区間、および雑音区間を検出する。例えば、送話信号について、短時間平均パワー(0.1～1s程度)と長時間平均パワー(1s～100s程度)を求め、短時間平均パワーと長時間平均パワーの比が送話の閾値を超える場合に送話区間と判定する。また、それぞれのマイクロホン受音信号について、短時間平均パワー(0.1～1s程度)と長時間平均パワー(1s～100s程度)を求め、短時間平均パワーと長時間平均パワーの比が受話の閾値を超えた場合に受話区間と判定し、雑音の閾値未満の場合に雑音区間と判定する。

【0118】

送受話判定部303の判定結果が送話区間であった場合に、音源位置検出部1

06は、第1の実施形態と同様にして音源位置を検出する。

【0119】

次に、共分散行列算出部107は、マイクロホン受音信号および受話信号の共分散を求め、それを行列にする。まず、マイクロホン受音信号の周波数領域変換信号を $X_1(\omega) \sim X_M(\omega)$ とし、受話信号の周波数領域変換信号を $Z(\omega)$ とする。これらの信号の共分散行列

【0120】

【外31】

$$R_{xx}(\omega)$$

【0121】

は式(20)により算出される。

【0122】

【数14】

$$R_{xx}(\omega) = \begin{pmatrix} Z(\omega) \\ X_1(\omega) \\ \vdots \\ X_M(\omega) \end{pmatrix} \begin{pmatrix} Z(\omega)^* & X_1(\omega)^* & \cdots & X_M(\omega)^* \end{pmatrix} \quad (20)$$

ただし、*は共役を表す。

【0123】

次に、共分散行列記憶部108では、音源位置検出部106の検出結果および送受話判定部303の判定結果に基づき、共分散行列

【0124】

【外32】

$$R_{xx}(\omega)$$

【0125】

を、各音源位置に対する共分散行列

【0126】

【外33】

$$\mathbf{R}_{\mathbf{S}_i \mathbf{S}_i}(\omega)$$

【0127】

と、送話区間の共分散行列

【0128】

【外34】

$$\mathbf{R}_{\mathbf{E}\mathbf{E}}(\omega)$$

【0129】

と、雑音区間の共分散行列

【0130】

【外35】

$$\mathbf{R}_{\mathbf{N}\mathbf{N}}(\omega)$$

【0131】

として保存する。

【0132】

収音レベル推定部109は、各音源の収音レベル P_{S_i} を、各音源位置に対する共分散行列

【0133】

【外36】

$$\mathbf{R}_{\mathbf{S}_1 \mathbf{S}_1}(\omega) \sim \mathbf{R}_{\mathbf{S}_n \mathbf{S}_n}(\omega)$$

【0134】

と、各音源位置ごとの任意のM+1行の行列

【0135】

【外37】

$$\mathbf{G}_1(\omega) \sim \mathbf{G}_n(\omega)$$

【0136】

に基づく式(21)により求める。行列

【0137】

【外38】

$$\mathbf{G}_i(\omega)$$

【0138】

は、例えば、1番目のマイクロホン1021の信号のみを取り出す行列

【0139】

【外39】

$$\mathbf{G}_i(\omega) = (1 \ 0 \ \dots \ 0)$$

【0140】

などである。

【0141】

【数15】

$$P_{Si} = \frac{1}{W} \sum_{\omega=0}^W \mathbf{G}_i(\omega)^H \mathbf{R}_{SiSi}(\omega) \mathbf{G}_i(\omega) \quad (21)$$

【0142】

次に、フィルタ係数算出部110では、各音源から発せられた音を所望の音量で収音するためのフィルタ係数を計算する。まず、各マイクロホン1021～102Mに接続されたフィルタ1031～103Mのフィルタ係数を周波数領域に変換したものを $H_1(\omega) \sim H_M(\omega)$ とし、受話信号をフィルタリングするフィルタ302のフィルタ係数を周波数領域に変換したものを $F(\omega)$ とする。次に、これらのフィルタ係数を式(22)により行列としたものを

【0143】

【外40】

 $\mathbf{H}(\omega)$

【0144】

とする。

【0145】

【数16】

$$\mathbf{H}(\omega) = \begin{pmatrix} F(\omega) \\ H_1(\omega) \\ \vdots \\ H_M(\omega) \end{pmatrix} \quad (22)$$

【0146】

また、受話区間のマイクロホン受音信号の周波数領域変換信号を $X_{E,1}(\omega) \sim X_{E,M}(\omega)$ 、受話信号の周波数領域変換信号を $Z_E(\omega)$ とし、雑音区間のマイクロホン受音信号の周波数領域変換信号を $X_{N,1}(\omega) \sim X_{N,M}(\omega)$ 、受話区間の受話信号の周波数領域変換信号を $Z_{NS}(\omega)$ とし、 i 番目の音源が発音している期间のマイクロホン受音信号の周波数領域変換信号を $X_{Si,1}(\omega) \sim X_{Si,M}(\omega)$ 、受話信号の周波数領域変換信号を $Z_{Si}(\omega)$ とする。

【0147】

ここで、フィルタ係数行列

【0148】

【外41】

 $\mathbf{H}(\omega)$

【0149】

に要求される条件は、マイクロホン受音信号および送話信号をフィルタ係数行列

【0150】

【外42】

 $H(\omega)$

【0151】

でそれぞれフィルタリングし、フィルタリング後の信号を加算したときに、エコ一信号と雑音信号が消去され、送話音声信号のみが所望のレベルで送話されることである。

【0152】

したがって、受話区間および雑音区間の信号に対しては、フィルタリングおよび加算後の信号が0となる式(23)および式(24)が理想条件となり、送話区間の信号に対しては、フィルタリングおよび加算後の信号がマイクロホン受音信号および受話信号に任意のM+1行の行列

【0153】

【外43】

 $G_r(\omega)$

【0154】

と所望の利得を乗じた信号となる式(25)が理想条件となる。行列

【0155】

【外44】

 $G(\omega)$

【0156】

は、例えば、1番目のマイクロホン1021の信号のみを取り出す行列

【0157】

【外45】

$$G(\omega) = (0 \ 1 \ 0 \ \cdots \ 0)$$

【0158】

などである。

【0159】

【数17】

$$\begin{pmatrix} Z_E(\omega) & X_{E,1}(\omega) & \cdots & X_{E,M}(\omega) \end{pmatrix} H(\omega) = 0 \quad (23)$$

$$\begin{pmatrix} Z_N(\omega) & X_{N,1}(\omega) & \cdots & X_{N,M}(\omega) \end{pmatrix} H(\omega) = 0 \quad (24)$$

$$\begin{pmatrix} Z_S(\omega) & X_{S,1}(\omega) & \cdots & X_{S,M}(\omega) \end{pmatrix} H(\omega) = \sqrt{\frac{P_{\text{out}}}{P_S}} \begin{pmatrix} Z_S(\omega) & X_{S,1}(\omega) & \cdots & X_{S,M}(\omega) \end{pmatrix} G_i(\omega) \quad (25)$$

【0160】

次に、式(23)～(25)の条件をフィルタ係数行列

【0161】

【外46】

$H(\omega)$

【0162】

について最小二乗解で解けば、式(26)となる。ただし、 C_E はエコー消去量に対する重みの定数であり、値が大きくなるほどエコーの消去量が増加する。その他の記号の意味については、第2の実施形態と同様である。

【0163】

【数18】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{S_i} R_{S_i S_i}(\omega) + C_N R_{NN}(\omega) + C_E R_{EE}(\omega) \right\}^{-1} \sum_{i=1}^n C_{S_i} \sqrt{\frac{P_{\text{out}}}{P_S}} R_{S_i S_i}(\omega) G_i(\omega) \quad (26)$$

【0164】

以上で、音量を調整し、雑音を抑圧するためのフィルタ係数を求める式(26)を導出した。

【0165】

次に、式(26)により求められたフィルタ係数はフィルタ部1031～103M, 302にコピーされ、マイクロホン受音信号と受話信号をそれぞれフィル

タリングする。フィルタリング後の信号は加算器104で加算され、可変増幅器111においてあらかじめ設定された固定増幅率で増幅された後、出力信号として出力される。

【0166】

また、式(26)を式(27)に置き換えてフィルタ係数を算出し、可変増幅器111の増幅率Bを、第1の実施形態と同様に最も新しく発音のあった音源の収音レベルP_{snew}から式(16)により設定するという方法もある。

【0167】

【数19】

$$H(\omega) = \left\{ \sum_{i=1}^n C_{Si} R_{Si}(\omega) + C_N R_{NN}(\omega) + C_E R_{EE}(\omega) \right\}^{-1} \sum_{i=1}^n C_{Si} R_{Si}(\omega) G_i(\omega) \quad (27)$$

【0168】

以上、第3の実施形態は、第2の実施形態における音量調整、雑音抑圧の効果に加え、エコー抑圧を行うことが可能である。

【0169】

これら以外の部分に関しては、第2の実施形態と同じであるので、説明を省略する。

【0170】

(第4の実施形態)

次に、本発明の第4の実施形態の収音装置について説明する。

【0171】

第4の実施形態の収音装置は、第1～3の実施形態の収音装置の各音源位置に対する感度拘束の重みの定数C_{Si}～C_{Sn}を時間的に変化させた装置である。

【0172】

時間的に変化する各音源位置に対する感度拘束の重みをC_{Si}(t)～C_{Sn}(t)とする。C_{Si}(t)～C_{Sn}(t)は、過去に発話された音源ほど小さくする。例えば、音源位置検出部106の検出結果よりi番目の音源から発音がある区間は、C_{Sj}(t)=1とし、i番目の音源から発音がない区間ではC_{Sj}(t)=q・C_{Sj}(t-1)として

、 i 番目の音源位置に対する感度拘束の重みを減衰させる。ただし、 q は、減衰の時定数を決定するための定数であり、 $0 \leq q \leq 1$ の値をとる。

【0173】

このように各音源位置に対する感度拘束の重みを変化させることにより、過去に発話のあった音源位置に対する感度拘束を減少させることができる。これにより、第1～3の実施形態の収音装置に比べ、感度拘束する音源数を抑えることができ、音量調整の性能および雑音、エコー抑圧性能を向上することができる。

【0174】

これら以外の部分に関しては、第1～3の実施形態と同じであるので、説明を省略する。

【0175】

(第5の実施形態)

本発明の第5の実施形態の収音装置について説明する。

【0176】

第5の実施形態の収音装置は、第1～4の実施形態の収音装置のフィルタ係数算出段階において、共分散行列

【0177】

【外47】

$$R_{xx}(\omega)$$

【0178】

を白色化することを特徴とする。

【0179】

白色化の方法は、共分散行列

【0180】

【外48】

$$R_{xx}(\omega)$$

【0181】

対角成分

【0182】

【外49】

$\text{diag}(\mathbf{R}_{xx}(\omega))$

【0183】

と任意のM行またはM+1行の行列

【0184】

【外50】

\mathbf{D}

【0185】

に基づく重み

【0186】

【外51】

$1/\{\mathbf{D}^H \text{diag}(\mathbf{R}_{xx}(\omega)) \mathbf{D}\}$

【0187】

を各々の共分散行列

【0188】

【外52】

$\mathbf{R}_{xx}(\omega)$

【0189】

に乗じることで共分散行列を白色化する。

【0190】

共分散行列の白色化により、フィルタ係数算出部110において求められるフィルタ係数は、送話信号、受話信号、雑音信号のスペクトルの変化に対して、変化しなくなる。これにより、スペクトル変化による音量調整性能、およびエコー

、雑音の抑圧性能の変化がなくなり、定常的な音量調整とエコー、雑音抑圧を得ることができる。

【0191】

これら以外の部分に関しては、第1～4の実施形態と同じであるので、説明を省略する。

【0192】

(第6の実施形態)

本発明の第6の実施形態の収音装置について説明する。

【0193】

第6の実施形態の収音装置は、第1～5の実施形態の収音装置の共分散行列記憶段階が、過去に記憶されている共分散行列と、共分散行列算出部107により新たに算出された共分散行列を平均して、現在の共分散行列として記憶することを特徴とする。

【0194】

共分散行列の平均は、例えば以下の方法により行われる。

【0195】

過去に記憶されている共分散行列を

【0196】

【外53】

$$\mathbf{R}_{\mathbf{XX},\text{old}}(\omega)$$

【0197】

とし、共分散行列算出部107により新たに算出された共分散行列を

【0198】

【外54】

$$\mathbf{R}_{\mathbf{XX},\text{new}}(\omega)$$

【0199】

とし、式(28)により平均された共分散行列

【0200】

【外55】

$$R_{xx}(\omega)$$

【0201】

を求める。ただし、 p は、平均の時定数を決定する定数であり、 $0 \leq p < 1$ の値をとる。

【0202】

【数20】

$$R_{xx}(\omega) = (1 - p) \cdot R_{xx,new}(\omega) + p \cdot R_{xx,old}(\omega) \quad (28)$$

【0203】

以上 の方法により、共分散行列を平均して記憶することにより、平均前より正確な共分散行列を求めることができる。

【0204】

これら以外の部分に関しては、第1～5の実施形態と同じであるので、説明を省略する。

【0205】

なお、本発明は専用のハードウェアにより実現されるもの以外に、その機能を実現するためのプログラムを、コンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行するものであってもよい。コンピュータ読み取り可能な記録媒体とは、フロッピーディスク、光磁気ディスク、CD-ROM等の記録媒体、コンピュータシステムに内蔵されるハードディスク装置等の記憶装置を指す。さらに、コンピュータ読み取り可能な記録媒体は、インターネットを介してプログラムを送信する場合のように、短時間の間、動的にプログラムを保持するもの（伝送媒体もしくは伝送波）、その場合のサーバとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含む。

【0206】**【発明の効果】**

以上説明したように、本発明によれば、複数のマイクロホンで受音した信号から音源位置を検出し、音源位置ごとの共分散行列および収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるようなフィルタ係数を算出し、これらのフィルタ係数で、マイクロホン受音信号をフィルタリングして加算することにより、話者が複数人存在し、それぞれの話者音声が異なるレベルで収音される場合においても音量を音源位置ごとに適切に調整した送話信号を得ることができる。

【図面の簡単な説明】**【図 1】**

本発明の第1の実施形態の収音装置を示すブロック図である。

【図 2】

本発明の第2の実施形態の収音装置を示すブロック図である。

【図 3】

本発明の第3の実施形態の収音装置を示すブロック図である。

【図 4】

本発明の音量調整効果を示す図である。

【図 5】

従来の収音装置を示すブロック図である。

【符号の説明】

1021～102M マイクロホン

1031～103M フィルタ

104 加算器

105 発話判定部

106 音源位置検出部

107 共分散行列算出部

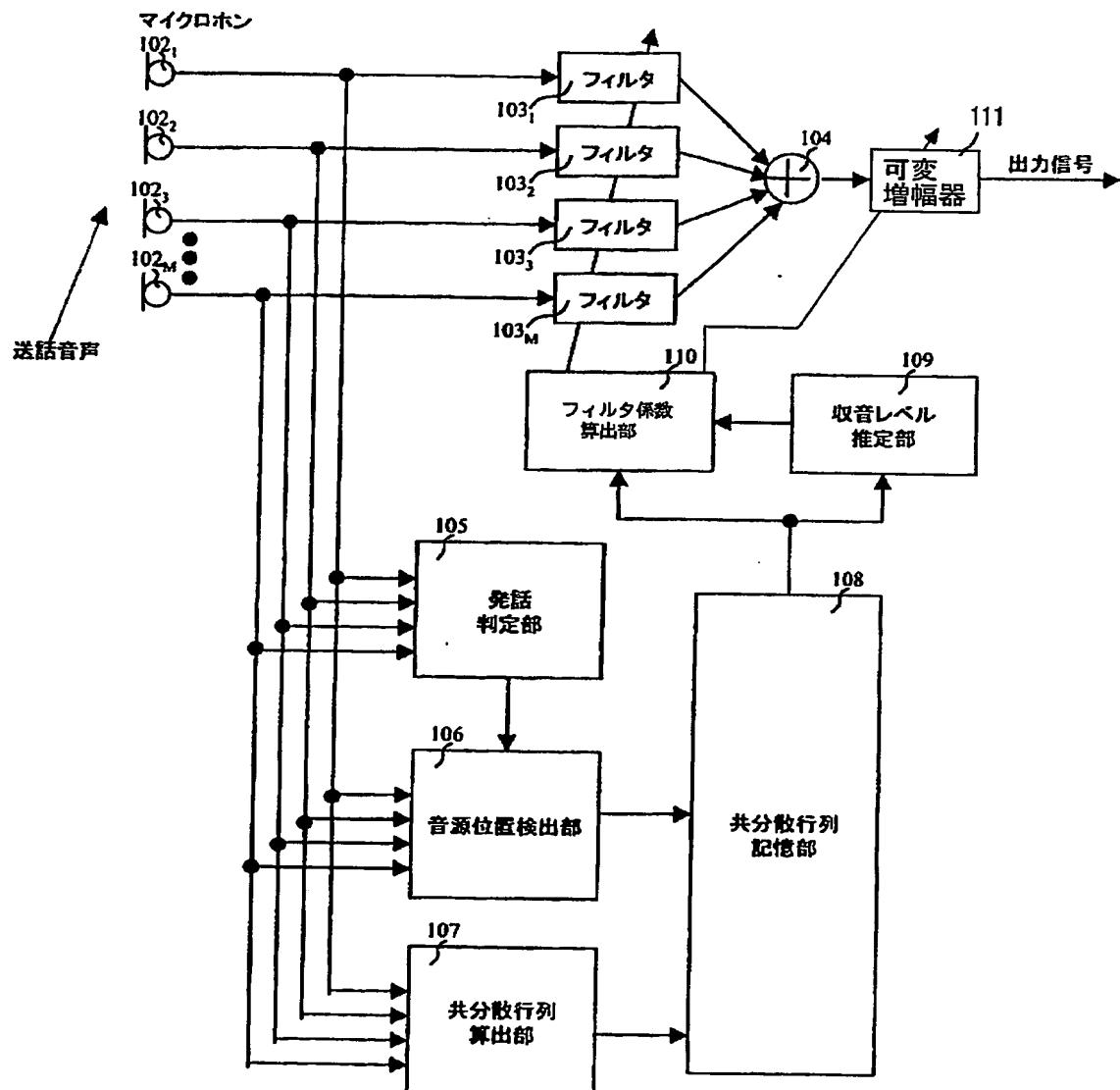
108 共分散行列記憶部

109 収音レベル推定部

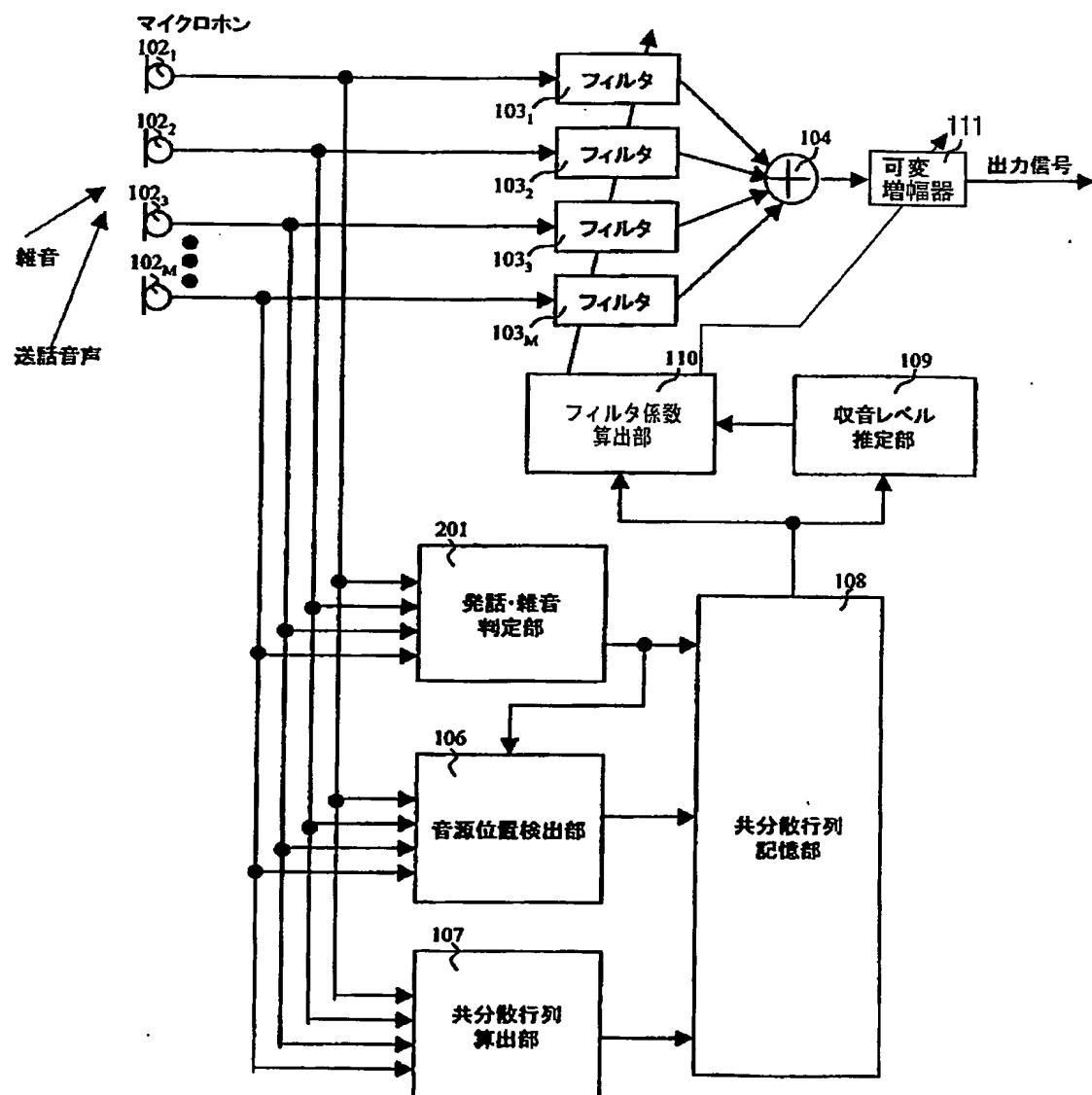
- 110 フィルタ係数算出部
- 111 可変増幅器
- 201 発話・雑音判定部
- 301 スピーカ
- 302 フィルタ
- 303 送受話判定部
- 401 マイクロホン
- 402 パワー算出部
- 403 増幅率設定部
- 404 増幅器

【書類名】 図面

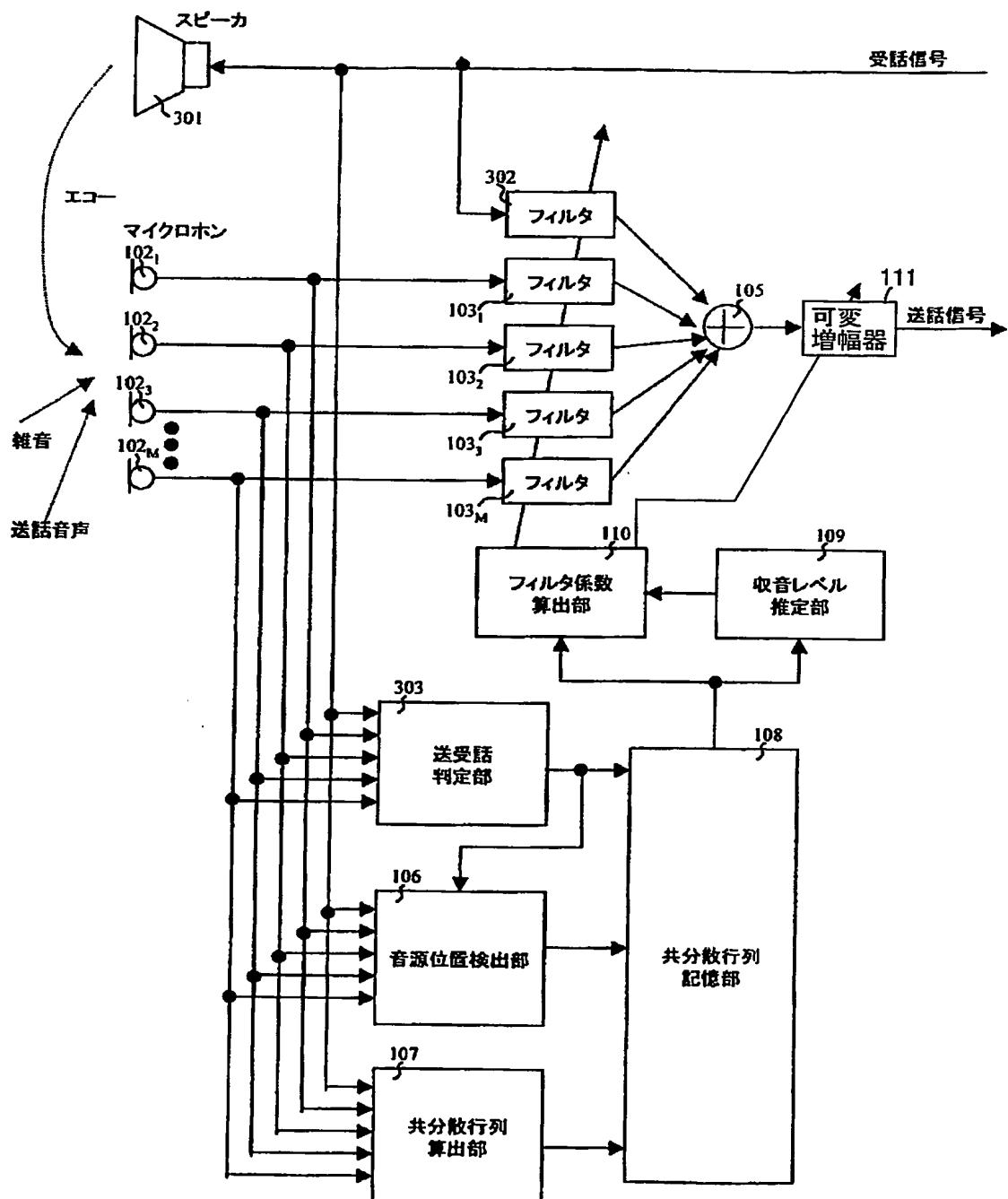
【図 1】



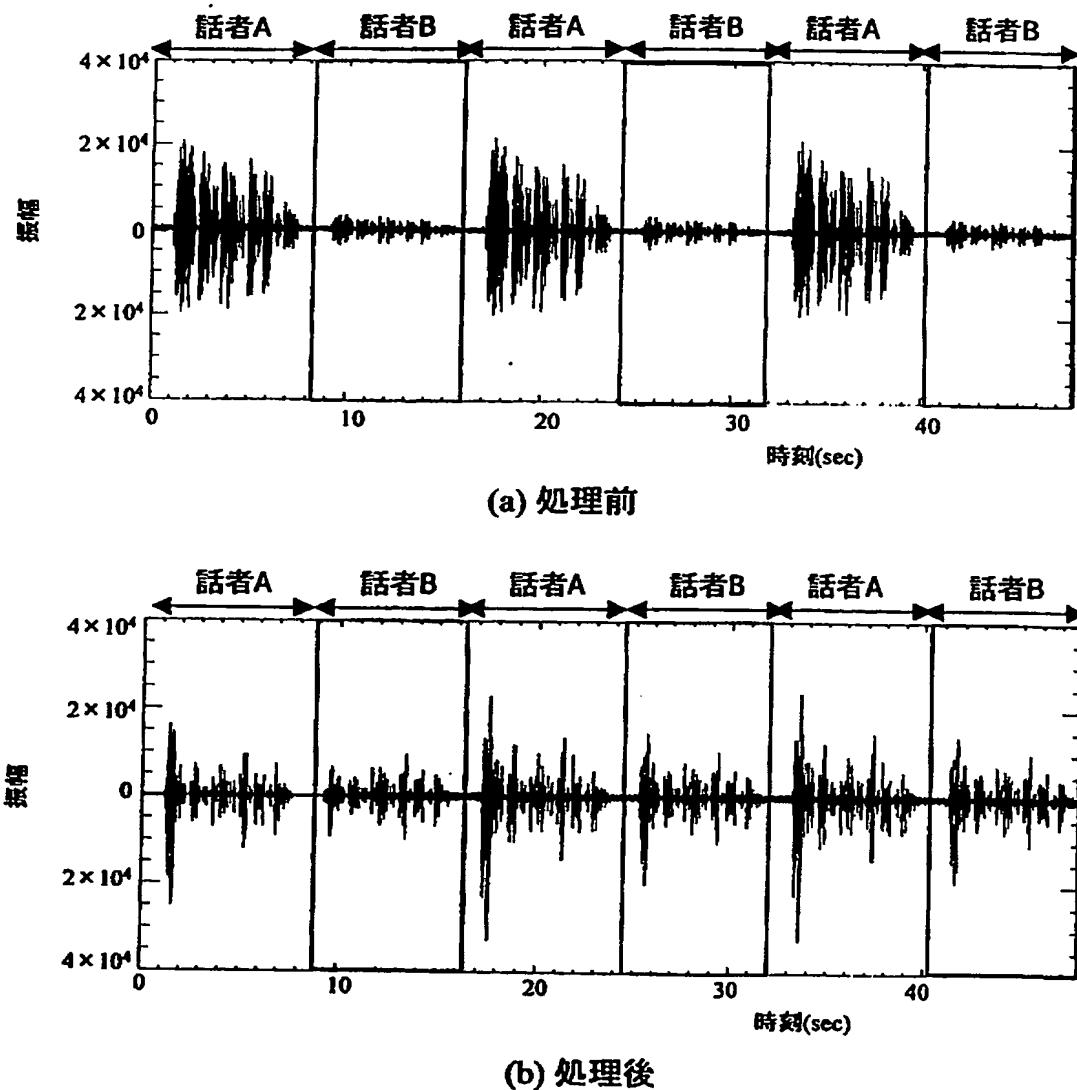
【図 2】



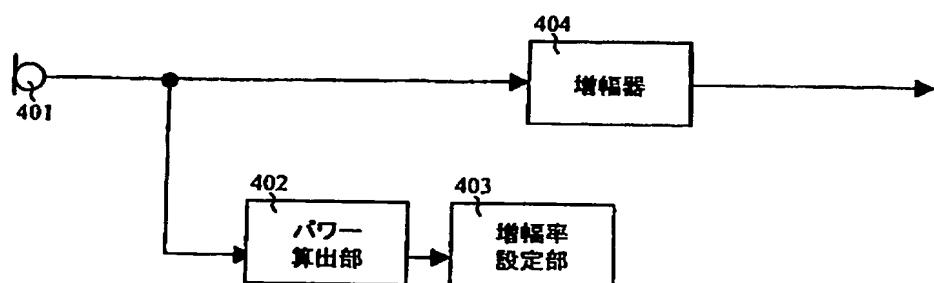
【図3】



【図4】



【図5】



【書類名】 要約書

【要約】

【課題】 話者が複数人存在し、それぞれの話者音声が異なるレベルで収音手段に収音される場合においても、各話者音声が適切な音量となるように音量を自動調整する。

【解決手段】 発話判定部105はマイクロホン102₁～102_Mの各々で受音されたマイクロホン受音信号から発話区間を判定する。音源位置検出部106は発話区間と判定された場合に、マイクロホン受音信号から音源位置を検出する。共分散行列算出部107はマイクロホン受音信号から共分散行列を算出する。共分散行列記憶部108は、共分散行列を音源位置ごとに記憶する。収音レベル推定部109は各音源位置に対する共分散行列に基づき、音源位置ごとの収音レベルを求める。フィルタ係数算出部110は前記共分散行列および前記音源位置ごとの収音レベルに基づく演算式を用いて、各音源の収音信号に対する送出レベルが各々所望の送出レベルとなるような複数チャネルのフィルタ係数と增幅率を算出する。

【選択図】 図1

出願人履歴情報

識別番号 [000004226]

1. 変更年月日 1999年 7月15日

[変更理由] 住所変更

住所 東京都千代田区大手町二丁目3番1号
氏名 日本電信電話株式会社

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.